

Hallucinated 3D Face Model from A Single 2D Low-Resolution Face Using Machine Learning

H.M.M Naleer¹ and Yao LU²

¹Faculty of Applied Sciences,

South Eastern University of Sri Lanka, Oluvil, Sri Lanka.

²Laboratory of Intelligent Information Technology,

Beijing Institute of Technology, Beijing 100081, China.

Corresponding author's e-mail: hmmnaleer@gmail.com vis vl@bit.edu.cn

Abstract

In this paper, the 3D face hallucination system is proposed on both 2D training face images as well as respective 3D training face models with grey-level. The proposed method hallucinates the 3D high-resolution model patch using same position of each image patches of interpolated 2D training image and 3D high-resolution training face model for low-resolution input image. Firstly, the optimal weights of the 2D input low-resolution image position-patches are estimated with the corresponding 2D low-resolution training image patches. The canonical correlation analysis (CCA) is used to learn the mapping between the 2D interpolated face training image and the 3D face model with respect to their weights. Secondly, the corresponding 3D face model patch with weight by matching high score among the 2D interpolated training image patches and 3D training face model is selected. Finally, the 3D high-resolution facial model is formed by integrating the hallucinated 3D patches which are obtained through mapping patches with respective weights. In order to evaluate the performances of the above approaches, we used example based learning methods to obtain the high-resolution output for a low-resolution input. In this approach, we used the available frontal data sets such as FERET, CAS-PERL and CMU to analysis the performance and some parameters are also considered, which may affect the results from the above proposed method.

Keywords: face hallucination, 3D face model, 2D low-resolution, CCA

Introduction

The modeling face is used in several computer vision applications, which are presented in (Banz, *et al.*, 2003, Edwards *et al.*, 1998, Banz & Vetter, 2003, Ramanathan, *et. al.*, 2006, Baker, *et al.*, 2004). The 3D models become useful universally by face enhancement of 3D models. In several aspects, the 3D face model super-resolution is employed. The capture of 3D data is limited according to the distance of the situation and objects. The availability of 3D scanners and capturing instruments are acquired typically low-resolution 3D data. Therefore, the hallucination of 3D face models is very helpful while the 3D data models are low-resolution. However, the time-consuming 3D scanner (Eisert & Girod 1998) is not accepted for such a face hallucination system. The method of using two orthogonal photos (Goto, *et al.*, 2002) cannot provide real-time processing either. The single image is used to obtain the super-resolution of face (Feng & Yuen 2000), where the head rotation parameters require measuring via other images. In addition, it gives the inaccurate results for 3D face models. On the other hand, the multiple images are used to obtain 3D super-resolution, is presented (Morris, *et al.*, 1999). This process is involved with 2D image and 3D model. The shape from stereo and shading are studied in (Morris, *et al.*, 1999, Zhang, *et al.*, 1995, Hom & Brooks, 1989) for 3D super-resolution. Though, few methods are presented for 3D super-resolution with an input 2D image. Presently, 3D data has become more advance while rapid process in 3D technology. This kind of 3D super-resolution is great needs in practice, since occasionally the 3D acquirement system does not provide good enough data due to 3D face recognition and lengthy distance (Gang & Zhaohui 2005).

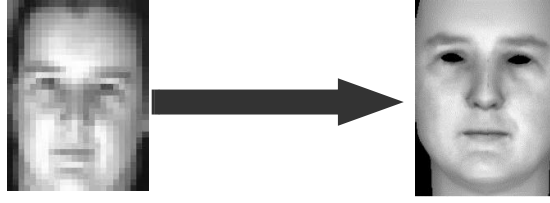


Figure 1.1: Input 2D low-resolution face image (left), Output 3D hallucinated face image (right)

The 3D position values of head shape are obtained by scanner technology which is the main approach to create 3D face model currently. It needs additional cost due to need of specific instrument. Meanwhile, the 3D model creation can be obtained from 2D images. Conversely, in this paper, a new method based on learning-method is focused to provide the 3D hallucinated face model for 2D low-resolution input image. Recently, very few works studied on 3D face hallucination. However, the super-resolution of 3D face is presented (Pan *et al.*, 2006), which is pursued by the work (Baker & Kanade 2000). Firstly, 3D shape is acquired by progressive resolution chain (PRC). Secondly, MAP approach is used to obtain the 3D face super-resolution (Yang *et al.*, 2007).

Furthermore, human may be used 2D image and it can be visualize into 3D image for better visualization that is bring new change for eye with brain role. Thus 3D image is may not be precise. If the 2D image is low-resolution, the human brain has very important role in acquiring the 3D image and it may not be precise. To overcome this problem, we proposed the hallucinated 3D face model from a 2D low-resolution input image. We can divide our whole system into two phases. We consider as *matching phase* between 2D interpolated training image and 3D training face model and *hallucination phase*. Output of proposed method is shown in Fig. 1.1. The rest of work is organized as follows: Section 3 is the fundamental idea of CCA. Section 4 describes our proposed hallucinated 3D face model via image patches and section 5 presents experimental results on construction of 3D hallucinated face model from a 2D low-resolution image. Finally, the conclusion is given in section 6.

1. Basic Idea of CCA

The dominant multivariate analysis is, Canonical Correlation Analysis (CCA) (Hardoon, *et al.*, 2004), which has been used in several applications such as post estimation and face matching (Melzer, *et al.*, 2003 & Naleer, *et al.*, 2013). To maximize the correlation between two sets, the CCA is constructing the subspace.

Given N pairs of sample (x_i, y_i) of (X, Y) , $i = 1, \dots, N$, where $X \in R^n$ and $Y \in R^n$. The mean of both X and Y is zero. The goal of CCA is to learn a pair of direction w_x and w_y to maximize the correlation between the two projections w_x^T and w_y^T , where T denotes the transpose, which is to maximize:

$$\rho = \frac{E[w_x^T X Y^T w_y]}{\sqrt{E[w_x^T X X^T w_x] E[w_y^T Y Y^T w_y]}}, \quad (3.1)$$

where $E[f(x, y)]$ - the experimental anticipation of the function $f(x, y)$. The covariance matrix of (X, Y) is

$$C(X, Y) = E \left[\begin{pmatrix} X \\ Y \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix}^T \right] = E \left[\begin{pmatrix} C_{XX} \\ C_{XY} \end{pmatrix} \begin{pmatrix} C_{YX} \\ C_{YY} \end{pmatrix}^T \right],$$

where C_{XX} and C_{YY} are within sets covariance matrices. Hence, ρ can be written as

$$\rho = \frac{w_x^T C_{xy} w_y}{\sqrt{w_x^T C_{xx} w_x w_y^T C_{yy} w_y}} \quad (3.2)$$

Let $A = \begin{pmatrix} 0 & C_{XY} \\ C_{YX} & 0 \end{pmatrix}$, $B = \begin{pmatrix} C_{XX} & 0 \\ 0 & C_{YY} \end{pmatrix}$, we can show that the solution $W = (w_x^T, w_y^T)^T$ amounts to the extreme points of the Rayleigh quotient:

$$r = \frac{W^T A W}{W^T B W} \quad (3.3)$$

The solution w_x and w_y can be obtained as solutions of the generalized eigen problem:

$$A W = B W \lambda \quad (3.4)$$

In case of the sample size is comparatively small; the CCA has a tendency to over-fit to the training image [17]. If $\lambda_x > 0$ and $\lambda_y > 0$ are regularization parameters, addition of $\lambda_x I$ and $\lambda_y I$ to C_{xx} and C_{yy} respectively, gives encouraging avoided outputs. Let $\lambda_x = \lambda_y = \lambda$, the object function of regularized CCA is to maximize ρ ,

$$\text{where } \rho = \frac{w_x^T C_{xy} w_y}{\sqrt{w_x^T (C_{xx} + \lambda I) w_x w_x^T (C_{yy} + \lambda I) w_y}}.$$

2. Hallucinated 3D Face Model via Image Patches

2.1 CCA Regression for 2D-3D Matching Phase

The following section used CCA to find out the results of mapping between 3D and 2D face models. In the learning phase, the 2D (interpolated 2D training dataset (24×32))-3D (3D training face model (96×128)) mapping is learnt from the training set which consists of K pairs of 2D-3D face image face model respectively. In the mapping phase, the most correlative 3D face shape to the 2D (2D low-resolution input image (24×32)) is found by 2D-3D mapping. In the learning stage, N pairs of 2D-3D face are given as $(X, Y) = [(x_k, y_k)]$, $(k = 1, 2, \dots, N)$, where (x_k, y_k) is a corresponding pair of 2D and 3D.

Furthermore, to reduce the computational complex, PCA is applied firstly to transform x_k and y_k into the lower dimensional spaces (known as *dimension reduction*). PCA transform matrices P_x and P_y are learnt from 2D and 3D training sets respectively. The PCA projections are computed as $X' = P_x^T (X - \bar{X})$ and $Y' = P_y^T (Y - \bar{Y})$, where \bar{X} and \bar{Y} are the mean faces of 2D face image and 3D face model respectively. On the other hand, two further projection directions w_x and w_y are learnt for linear CCA by performing CCA on X' and Y' respectively (known as CCA regression). Also, let $w_x^T X'$ and $w_y^T Y'$ are best correlated. In the mapping phase, also need to project new pair of images X_{ts} and Y_{ts} into PCA firstly, that is $X'_{ts} = P_x^T (X_{ts} - \bar{X})$ and $Y'_{ts} = P_y^T (Y_{ts} - \bar{Y})$ respectively. For linear CCA, X'_{ts} and Y'_{ts} are projected into CCA sup-space, that is

$$X_{out} = w_x^T X'_{ts}, Y_{out} = w_y^T Y'_{ts} \quad (4.1)$$

Finally, the matching score (So) can be calculated as

$$So(X_{out}, Y_{out}) = \frac{X_{out} \cdot Y_{out}}{\|X_{out}\| \|Y_{out}\|} \quad (4.2)$$

2.2 3D Face Model Hallucination via Learning Method

In this section, the two phases as mapping phase and hallucination phase are measured. As mentioned in section 4.1, the 3D face model patch corresponding to 2D input low-resolution image patch can be found. The description of weights calculation on 2D low-resolution training image patches with input 2D low-resolution image is described in this section firstly. Secondly, hallucination phase is considered.

4.2.1 Calculation of Mapping Scores

We adopted the CCA to get the exact model of 2D-3D face mapping by patch based CCA, the following steps are considered in mapping stage.

- A large number of patches are created from 2D face image and 3D face models.
- CCA is used to learn the performance between 2D image and 3D face model for every patch.
- In the mapping phase, the testing images are also departed into several parts in the same way. Then we can get a score for each patch.
- The final score is obtained by combining the individual scores of all participating patches.
- Finally, selected the 3D patches with weights on 3D training face model set using mapping score to the corresponding 2D patches on interpolated 2D training set.

The weights are calculated as follows and the style of the patch is shown in Fig. 4.1. In the patch based weights in the particular location (i, j) is calculated by the following Eq. (4.3), which is described in (Naleer, *et al.*, 2013) Algorithm-I, where S - 2D column vector of low-resolution training images $[I_L^q]_{q=1}^n$. If (i, j) is indicated as coordinate of the image patch in the particular location, the weight is calculated by Eq. (4.3)

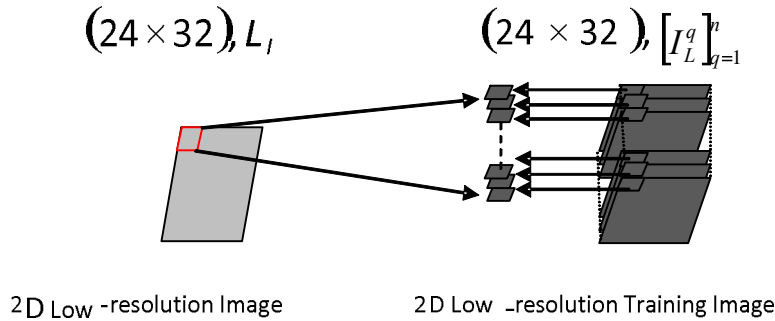


Figure 4.1: Weights calculation with respective patches

$$w_i(i, j) = \frac{[(L_i(i, j))_{i=1}^P \times R^T - (I_L^i(i, j))_{i=1}^P]^{-1} \times [(L_i(i, j))_{i=1}^P \times R^T - (I_L^i(i, j))_{i=1}^P]}{R^T \times \{[(L_i(i, j))_{i=1}^P \times R^T - (I_L^i(i, j))_{i=1}^P]^T \times [(L_i(i, j))_{i=1}^P \times R^T - (I_L^i(i, j))_{i=1}^P]\}^{-1} \times R} \quad (4.3)$$

The selection of patch size is an important issue. It is too large or it is too small, in both cases the local mapping of whole face information will be difficult. Hence the scores of each patch are combined to such a decisive level so that it is simplified. Various combining schemes include Median Rule and, Sum Rule, Majority Voting, Max Rule, Product Rule and Min Rule [6]. Eq. (4.4) gives the results of final matching score where, So_i and w_i are the output score and corresponding weight for i^{th} patch respectively.

$$F = \sum_{i=1}^P w_i \times So_i \quad (4.4)$$

4.2.2. 3D Hallucination Phase

Once we found the 2D face training image patch position with respective weights to corresponding 2D low-resolution image patch in the same position, according to the final matching score using CCA regression mapping between 3D training face model and 2D interpolated training images using weights, the following process are used to construct the final output as hallucinated 3D face model. According to the original position, the final 3D hallucinated face model is obtained by combining the 3D face model patches. Average of the pixels values in the overlapping areas among two adjacent hallucinated patches is used to acquire the final result for pixel of the overlapping areas. The proposed 3D face hallucination model has been summarized below and the entire framework is given in Fig. 4.2.

Step1: Find the weights between 2D low-resolution input and 2D low-resolution training images in the same positions using Eq. (4.3).

Step2: Find the fusion score between 2D training image and 3D face image model using Eq. (4.4).

Step3: Once selected the 3D face mode patch with respective weights according to the fusion score, the hallucinated 3D face model I_H^P can be obtained by following Eq. (4.5).

$$I_H^P(i, j) \cong \sum_{q=1}^n (I_H^q(i, j)) w_i(i, j) \quad (4.5)$$

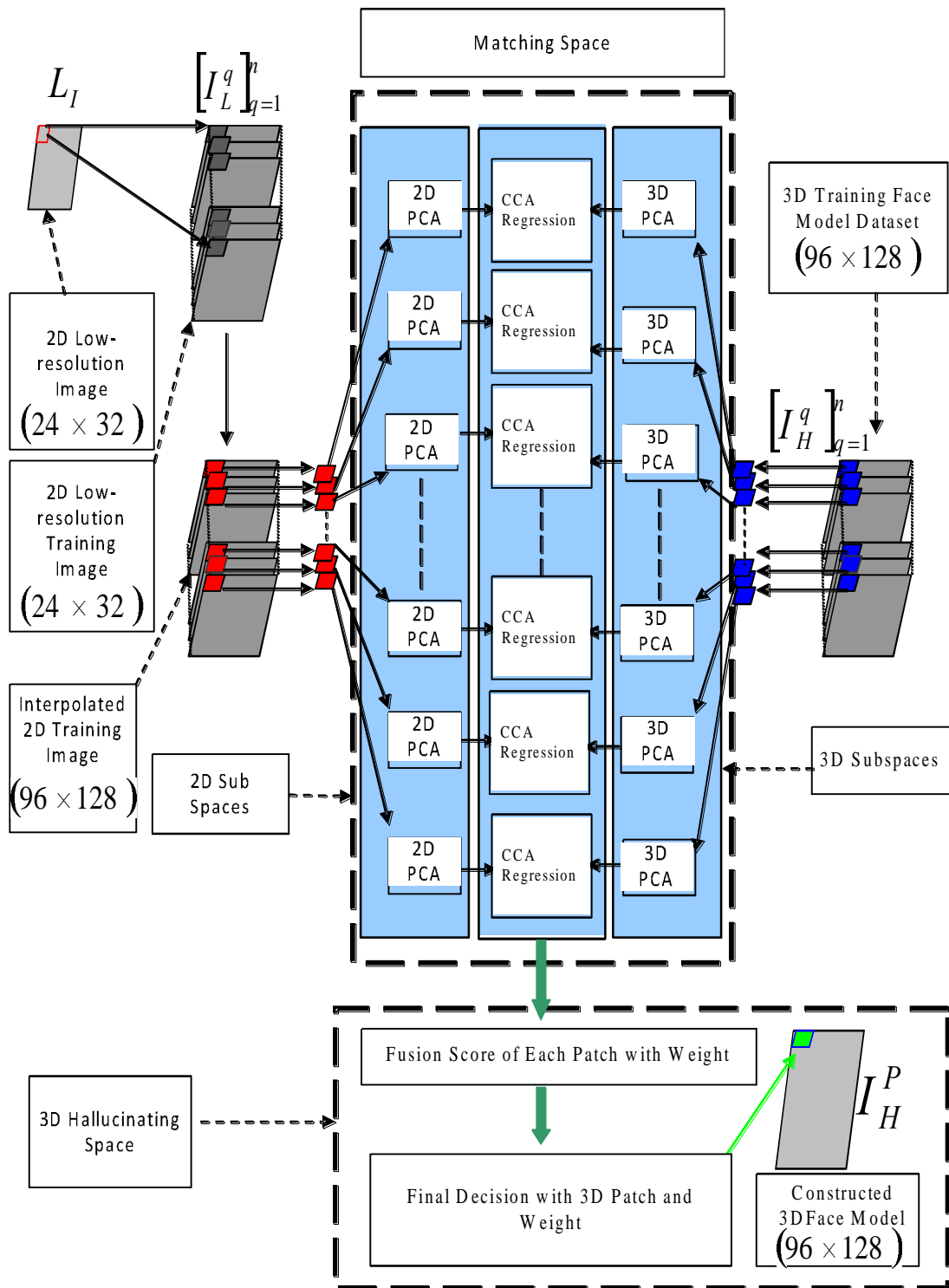


Figure 4.2: Framework of CCA based hallucinated 3D face Model

Experiments and Results

Our experiments are conducted with the CAS-PEAL and FERET face datasets consisting of 200 individuals. The training face data sets are formed as 2D face images and 3D face models pairs for respective each individual. The laser scanner is used to obtain 3D training face model, which is offer range images and the faces are essentially 2.7D data. In the 3D face model training set, have a little facial expression and head pose varying for corresponding 2D training data set of each individual. All 3D training face models (96×128) and 2D low-resolution face images (24×32), and all training images are manually aligned based on eyeballs, mouths positions with centers of left and right eyeballs and center of the mouth, and the input low-resolution image is also manually aligned. Our proposed approach is evaluated on 80 test low-resolution inputs. Some output of the test images on CASE-PERL are presented in Fig. 5.1.

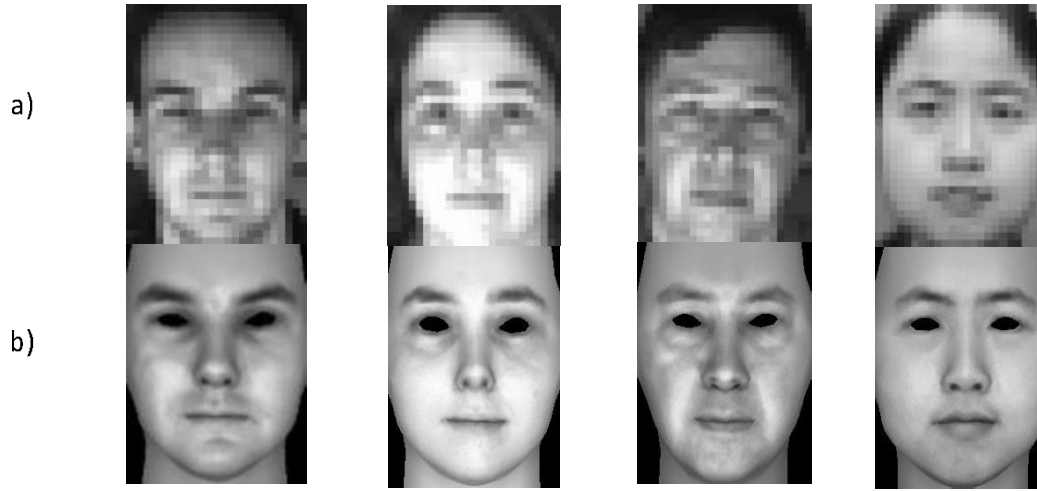


Figure 5.1: Output of proposed system: a) Input 2-D low-resolution image (24×32), b) Output 3D hallucinated image (96×128)

The comparative results for input low-resolution face image for each patch (3×3) are given in Fig. 5.2 in terms of cumulative match curves. Meanwhile the performance with training samples is also considered the results by cubic interpolation method and it is shown in Fig 5.3. To exhibit the performance of our proposed method, the amount of training images with RMS values has been given in Fig. 5.3. In fact, the cubic interpolation is not depending on the number of training images. Meanwhile, the performance of the proposed method is very poor while less than about 73 training images. It means, our methods is enclosed most of the variance of the 3D face model.

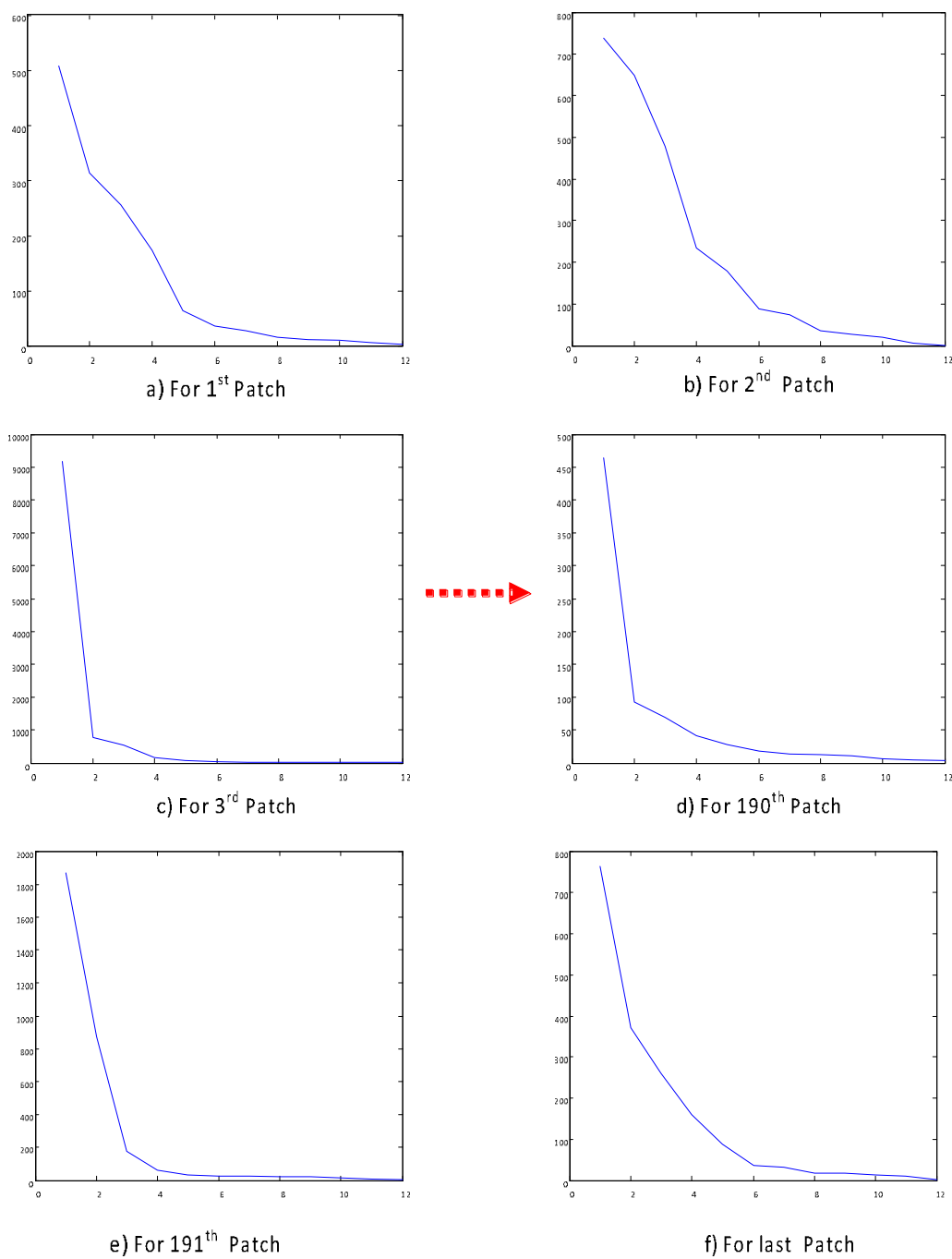


Figure 5.2: Cumulative match curve for 2D face image and 3D face model for each input low-resolution patch.

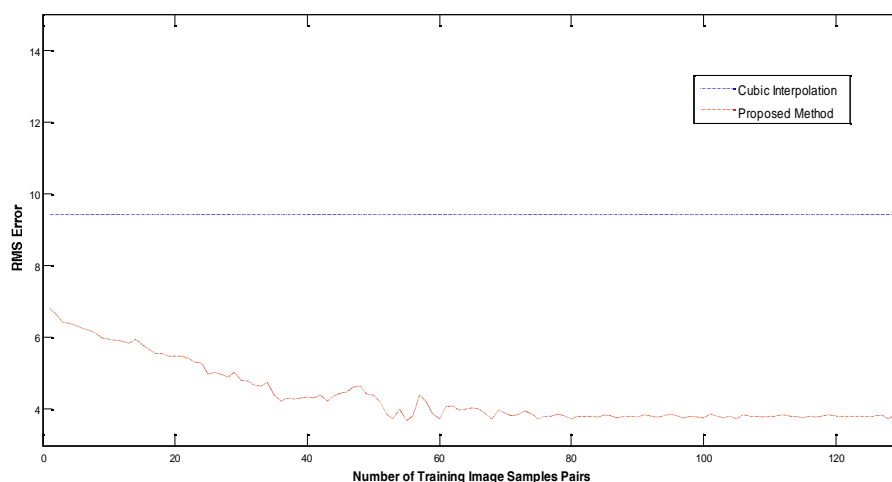


Fig. 5.3: Performance vs number of image sample pairs.

Conclusion

This work has presented a learning-based 3D face hallucination method for single 2D low-resolution face image with 2D low-resolution face image and 3D face model training image pairs based on matching scores. The experiments show that, it could be applicable to generate a 3D hallucinated face image version of the 2D low-resolution input face image in absence of 3D high-resolution face model in the training sets.

References

- Blanz, V, Basso, C, Poggio, T, & Vetter, T 2003, 'Reanimating Faces in Images and Video', *In Computer Graphics Forum*, vol. 22(3), pp. 641-650.
- Edwards, G, Cootes T, & Taylor, C1998, 'Face Recognition Using Active Appearance Models', *European Conference on Computer Vision*, vol. II (ECCV) [C], pp. 581-595.
- Blanz, V, & Vetter, T 2003, 'Face Recognition Based on Fitting A 3d Morphable Model', *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 25(9), pp. 1063-1074.
- Ramanathan, S, Kassim, A, Venkatesh ,Y & Wah, W 2006, 'Human Facial Expression Recognition Using A 3D Morphable Model', *Inter. Conf. on Image Processing (ICIP) [C]*, pp. 661-664.
- Baker, S, Matthews, I & Schneider, J 2004, 'Automatic Construction of Active Appearance Models as an Image Coding Problem', *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 26 (10), pp. 1380-1384.
- Eisert, P & Girod, B 1998, 'Analyzing Facial Expressions for Visual Conferencing', *IEEE Computer Graphics and Application*, vol. 18(5), pp. 70-78.
- Goto, T, Lee, W & Thalmann, N 2002, 'Facial Feature Extraction For Quick 3D Face Modeling', *Signal Process and Image Communication*, vol. 17(3), pp. 243-259.
- Feng, GC & Yuen, PC 2000, 'Recognition of Head-&-Shoulder Face Image Using Virtual Frontal-View Image', *IEEE Trans. Systems Man and Cybernetics Part A*, vol. 30(6), pp. 871-883.
- Morris, RD, Cheeseman, P, Smelyanskiy, VN & Maluf, DA 1999, 'A Bayesian Approach to High Resolution 3D Surface Reconstruction from Multiple Images', *IEEE Signal Processing Workshop [C]*, pp. 140-143.
- Zhang, Z, Deriche, R, Faugeras, O & Luong, QT 1995, 'A Robust Technique For Matching Two Uncalibrated Images Through The Recovery Of The Unknown Epipolar Geometry', *Artificial Intelligent*, vol. 78(1), pp. 87-119.

- Horn, B & Brooks, M 1989, 'Shape from Shading', *MIT Press*, pp. 123-171.
- Gang, P & Zhaohui, W 2005, '3D Face Recognition from Range Data', *Inter. Journal of Image and Graphics*, vol. 5(3) pp. 1-21.
- Pan, G, Han, S, Wu, Z & Wang, Y 2006, 'Super-resolution of 3D Face', *In Proc. European Conf. on Computer Vision (ECCV) [C]*, vol. 3952, pp. 389-401.
- Baker, S & Kanade, T 2000, 'Hallucinating Faces', *In Proc. Inter. Conf. on Automatic Face and Gesture Recognition. (FG) [C]*, pp. 83-88.
- Yang, Q, Yang, R, Davis, J & Nister, D 2007, 'Spatial-Depth Super Resolution for Range Images', *In Proc. IEEE Conf. Computer Vision Pattern Recognition (CVPR) [C]*, pp. 1-8.
- Hardoon, DR, Szedmak SR, & Shawe-taylor JR 2004, 'Canonical correlation analysis: An Overview with Application to Learning Methods', *Neural Comput*, vol. 16(12), pp. 2639– 2664.
- Melzer, T, Reiter, M & Bischof H 2003, 'Appearance Models Based on Kernel Canonical Correlation Analysis', *Pattern Recognition*, vol. 36(9), pp. 1961–1971.
- Naleer HMM, Yao Lu, & Feng Lv 2013, 'Effective Two-Step Method for Face Hallucination Based on Sparse Compensation on Over Complete Patches', *IET image processing*, vol. 7(2), pp. 624-632,